

Introducción. **Modelos de datos**

Se llama **datos** al conjunto de propiedades que caracterizan un fenómeno, e **información** al conjunto de valores que pueden tomar estas propiedades junto con las relaciones o dependencias entre las mismas.

Un **modelo de datos** es una representación abstracta de un determinado fenómeno real o abstracto. Podemos distinguir dos partes en un modelo de datos: la *estructura*, que es la representación abstracta del problema, y la *información*, que son los datos con que se alimenta el modelo. La medición de los valores correspondientes a los datos se realiza haciendo uso de unos tipos o clases abstractas de datos básicos (entero, real,...) que toman valores de un conjunto predefinido de antemano (*dominio*).

El primer paso en la representación de un problema del mundo real es la caracterización del mismo, o lo que es lo mismo, la determinación mediante un proceso de simplificación de los datos de interés (de entre todos los que intervienen en el problema) y sus límites (*universo del discurso*).

La *abstracción* es la capacidad mediante la cual una serie de objetos con propiedades comunes se categorizan en un nuevo objeto mediante una función de pertenencia. Los modelos de datos ofrecen distintos niveles de abstracción que facilitan la representación de los datos:

1. **Clasificación.** Es la acción de crear una categoría a partir de las características comunes a un conjunto de ejemplares.

Por ejemplo, a partir de los elementos Pedro, Juan y Cristina podemos crear la categoría profesor de instituto.

Su proceso inverso es la *particularización*.

2. **Agregación.** Es la capacidad de considerar un objeto en base a los elementos que lo constituyen.

Por ejemplo, podemos crear la clase coche a partir de las clases volante, ruedas, motor y carrocería.

Su proceso inverso es la *desagregación*.

3. **Generalización.** Es similar a la clasificación, pero creando una categoría a partir de las características comunes a un conjunto de otras categorías.

Por ejemplo, a partir de las categorías profesor de matemáticas, profesor de física y profesor de informática, podemos crear la categoría profesor de instituto.

Su proceso inverso es la *especialización*.

Podemos hablar de tres tipos de modelos de datos:

- **Modelo conceptual.** Describe los tipos o clases de objetos desde un punto de vista estructural. Para cada uno de estos tipos de objetos describe sus propiedades y el dominio y restricciones de cada una, así como las relaciones entre ellos. (Modelo E-R).
- **Modelo lógico.** Representa el problema bajo las restricciones específicas del tipo de SGBD que se aplique en cada caso específico. (Modelo Relacional para el caso de los SGBD Relacionales).
- **Modelo físico.** Representa el problema desde el punto de vista de su implementación en el sistema de tratamiento utilizado (SGBD concreto) y los métodos y mecanismos que se van a usar en su almacenamiento.

Tema 1. Modelo Entidad – Relación (E–R)

Propuesto por Chen en dos artículos (1976 y 1977). Es un modelo muy extendido que ha experimentado una serie de ampliaciones a lo largo de los años (no existe un modelo único). Sus características actuales permiten la representación de cualquier tipo de sistema y a cualquier nivel de abstracción, ofreciendo al diseñador una herramienta para aislar al modelo de consideraciones relativas a la máquina y a los usuarios.

Conceptos básicos del modelo E–R:

1. **Entidad:** es un objeto definido en base a la agregación de una serie de atributos. Corresponde a la caracterización de objetos del mundo real.

Se llama *tipo de entidad (intensión de entidad)* al tipo de objeto abstracto y *entidad* a cada uno de los ejemplares de ese tipo, aunque normalmente se simplifica y se llama entidad al tipo abstracto. La *extensión de entidad* es el conjunto de ejemplares que forman parte del tipo de entidad en un momento dado.

Se representa por un rectángulo conteniendo el nombre del tipo de entidad en letras mayúsculas.

Los tipos de entidad se pueden clasificar en:

- **Regulares:** sus ejemplares tienen existencia en sí mismos. Por ejemplo: LIBRO.

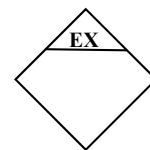


- **Débiles:** la existencia de los ejemplares de estos tipos de entidad está condicionada a la existencia de ejemplares de otro tipo de entidad. Por ejemplo, un ejemplar del tipo de entidad EDICIÓN no existiría si no hubiera un ejemplar correspondiente en el tipo de entidad LIBRO. Se representan por dos rectángulos, uno dentro del otro.

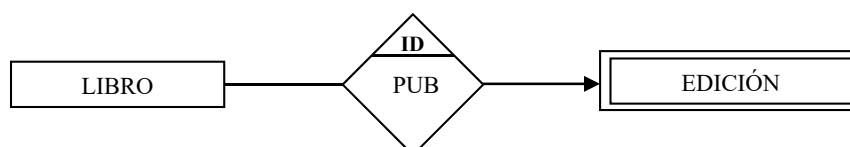
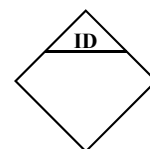


Hay dos tipos de dependencia de los tipos de entidad débiles respecto a los regulares:

Dependencia en existencia: los ejemplares del tipo de entidad débil no pueden existir si desaparece el ejemplar del tipo de entidad regular con el que están relacionados, pero pueden ser identificados sin necesidad de identificar el ejemplar del tipo de entidad fuerte. Se representa:



Por identificación: los ejemplares del tipo de entidad débil no tienen sentido en sí mismos y no pueden ser identificados sin identificar el ejemplar del tipo de entidad fuerte relacionado. La interrelación con dependencia en identificación se representa:

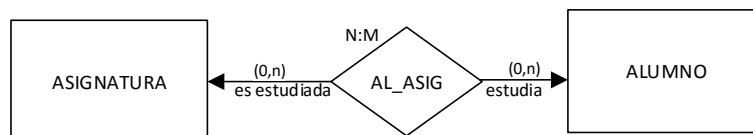


2. **Interrelación:** representa la relación existente entre entidades. La *intensión* o *tipo de interrelación* es la estructura genérica de vinculación entre los tipos de entidad, y la *extensión* o *conjunto de interrelaciones* es el conjunto de todas las relaciones entre los elementos de los tipos de entidad vinculados.

Se representa por un rombo etiquetado con el nombre del tipo de interrelación en mayúsculas unido mediante arcos a los tipos de entidad que vincula.

Los elementos y características de un tipo de interrelación son:

- **Nombre.** A veces se suele usar como nombre partes de los nombres de los tipos de entidad que relaciona.
- **Grado.** Es el número de tipos de entidad que participan en la interrelación. La mayoría de las interrelaciones suelen ser binarias (grado 2). Las interrelaciones de grado superior a 2 deben ser sustituidas, siempre que sea posible, por otras de grado 2.
- **Tipo de correspondencia.** Es el número máximo de ejemplares de un tipo de entidad que pueden estar relacionados con un ejemplar de otro tipo de entidad. Se representa, en el caso de interrelaciones de grado 2, con las cardinalidades máximas de cada uno de los tipos de entidad, en la forma 1:1, 1:N ó N:M, y poniendo una flecha (notación Bachman) en el arco que une los tipos de entidad con cardinalidad máxima indefinida.
- **Papel o rol.** Indica la forma en que participa cada tipo de entidad. Se indica en el arco que une el tipo de interrelación con el tipo de entidad. Puede omitirse cuando es evidente.
- Además, para cada interrelación es necesario representar la *cardinalidad* con que interviene cada tipo de entidad en el tipo de interrelación. Se representa etiquetando el arco mediante una pareja de datos de la forma (cardinalidad_mínima, cardinalidad_máxima).



Al igual que con los tipos de entidad, podemos distinguir una clasificación de los tipos de interrelación: tipos de interrelación *regulares* y tipos de interrelaciones *débiles*, en función de si relacionan tipos de entidad regulares entre sí o con tipos de entidad débiles.

Entre dos tipos de entidad puede existir más de un tipo de interrelación.

3. **Dominio:** representa el conjunto de valores posibles de una determinada propiedad o característica de un tipo de entidad o de un tipo de interrelación. En términos de abstracción, es una especialización de un conjunto.

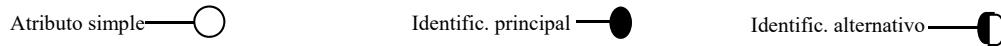
Se representa por un círculo pequeño acompañado de su nombre en minúsculas.

4. **Atributo:** es cada una de las posibles propiedades o características de un tipo de entidad o interrelación.

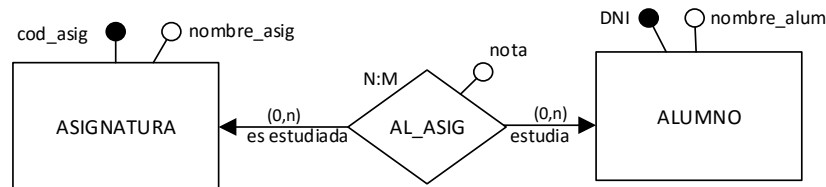
Se representa por un arco que une el tipo de entidad al que está asociado con el dominio en que toma valores el atributo (dominio asociado). Normalmente se simplifica la

representación del dominio y el atributo, y se representan ambos por un círculo pequeño unido por un arco al tipo de entidad y acompañado del nombre del atributo.

Además, según sea el símbolo podemos tener:

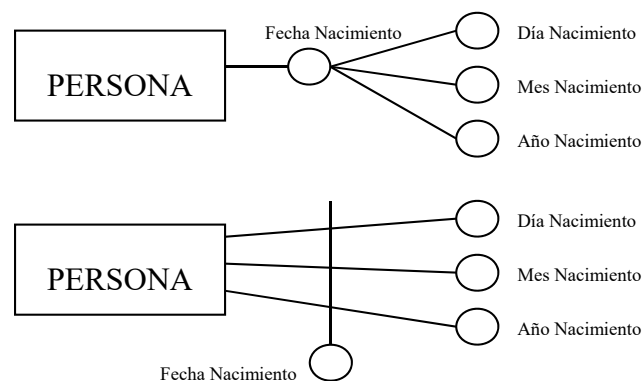


El *identificador candidato* de un tipo de entidad es el atributo o conjunto mínimo de atributos que no toma el mismo valor para dos entidades distintas de ese tipo. Si hay más de un conjunto de atributos que satisfagan esta condición, a uno de estos grupos



se le asignará el papel de *identificador principal* y al resto el de *identificador alternativo*. Cada tipo de entidad debe tener al menos un identificador principal.

Dentro del modelo E–R también pueden existir *atributos compuestos*, que se pueden representar de una las siguientes formas:



El modelo E–R también admite atributos *multivaluados* (en cuyo caso podríamos hablar también de cardinalidades mínima y máxima del atributo) y *opcionales*.

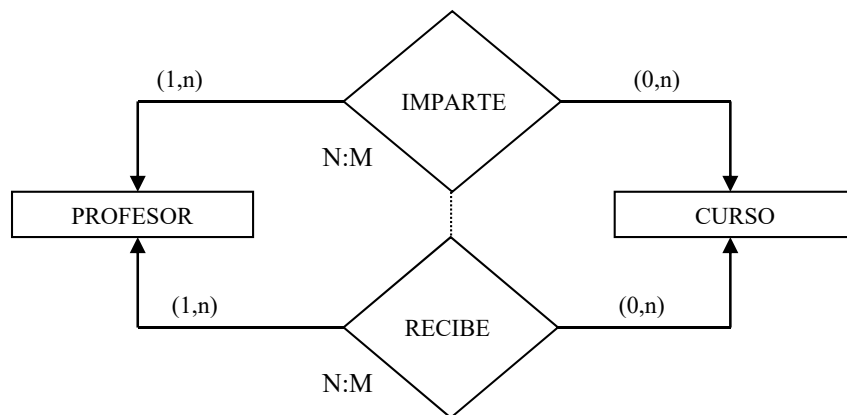


En el modelo E–R podemos tener los siguientes tipos especiales de interrelación:

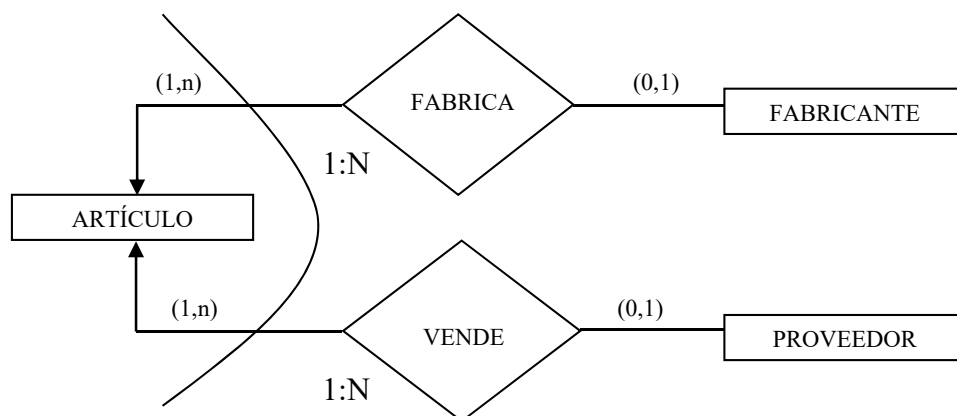
1. *Interrelaciones reflexivas*. Son tipos de interrelación en las que interviene un único tipo de entidad (unarias).



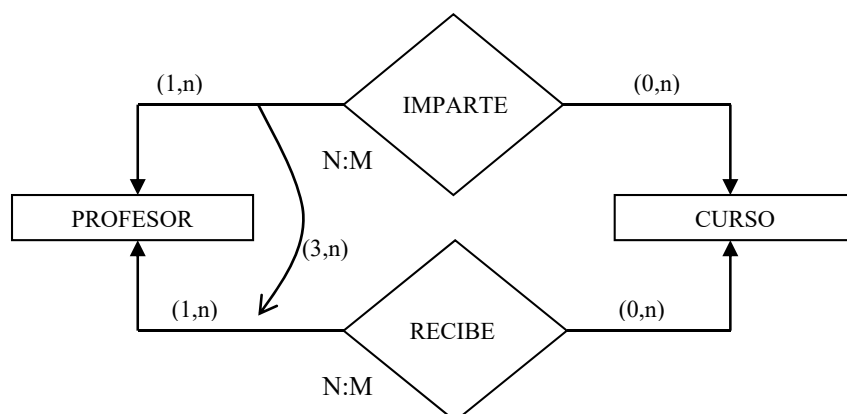
2. *Interrelaciones con restricciones de exclusividad.* Dos tipos de interrelación que implican a un mismo tipo de entidad participan de una restricción de exclusividad si los ejemplares de ese tipo de entidad pueden participar de uno u otro tipo de interrelación, pero no de ambos.



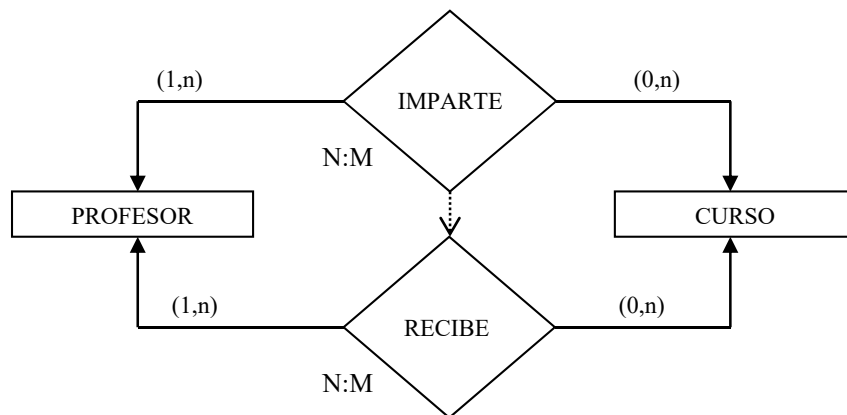
3. *Interrelaciones exclusivas.* Dos tipos de interrelación entre los mismos dos tipos de entidad son exclusivas si un ejemplar del primer tipo de entidad y otro ejemplar del segundo tipo de entidad sólo pueden estar relacionados por uno de las dos tipos de interrelación, nunca por ambos simultáneamente.



4. *Interrelaciones con restricciones de inclusividad.* Son dos tipos de interrelación que implican a un mismo tipo de entidad, en los que las entidades tienen que haber participado de un tipo de interrelación con una cardinalidad determinada para poder participar del otro.



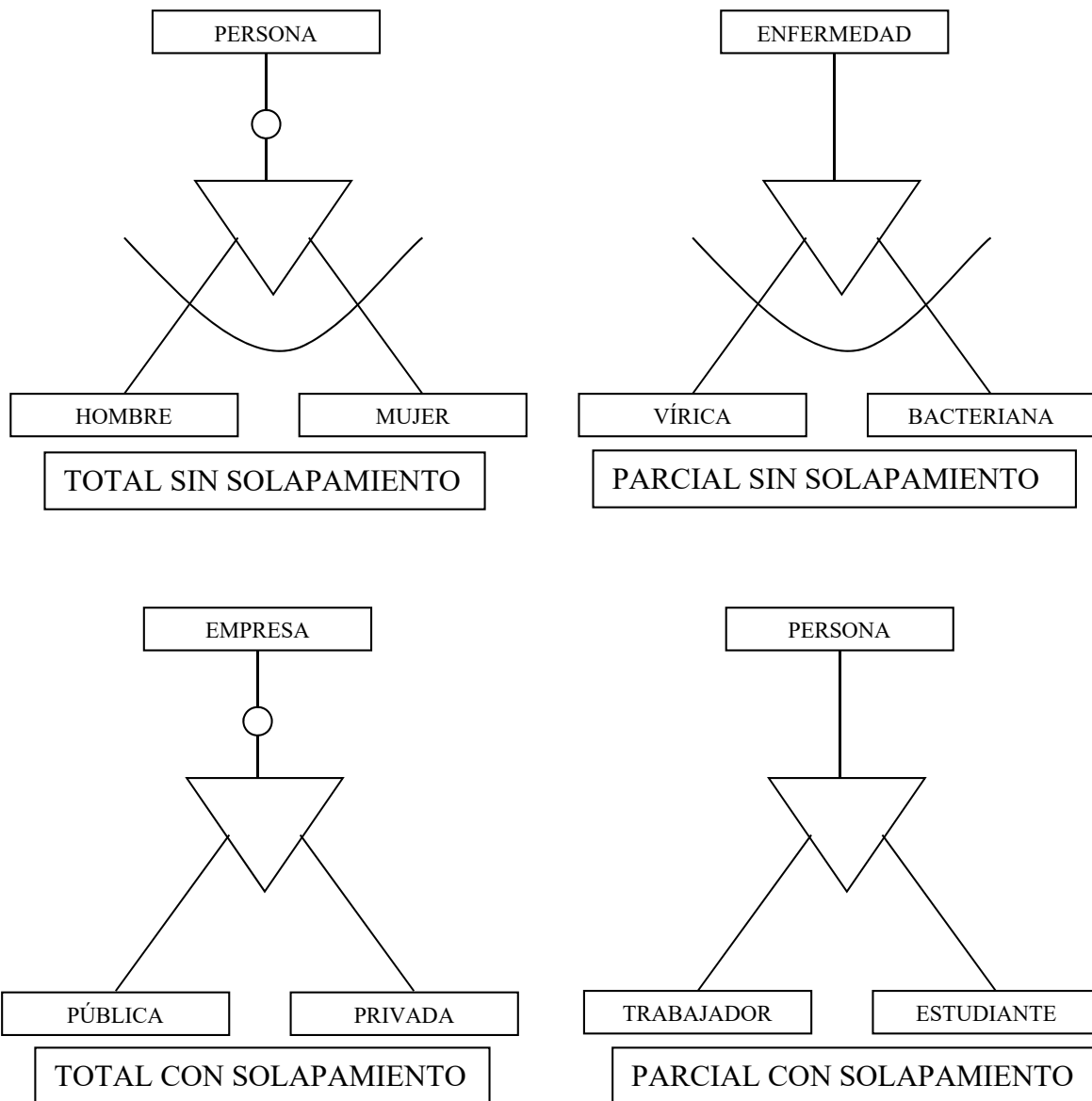
5. *Interrelaciones inclusivas*. Son aquellas que se establecen entre los mismos dos tipos de entidad y que restringen una interrelación entre dos entidades de cada uno de los tipos de entidad a la vinculación de esas dos mismas entidades a través del otro tipo de interrelación.



6. *Interrelaciones jerárquicas de generalización – especialización*. Se pueden clasificar a su vez en:

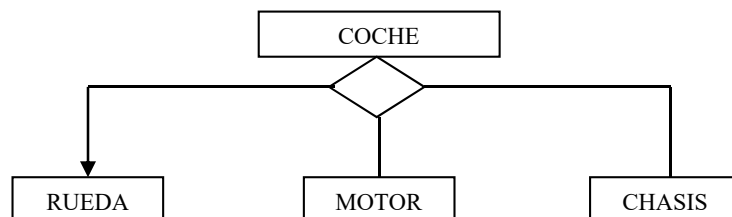
- *Parciales* o *totales*, dependiendo de si pueden existir o no ejemplares del tipo de entidad general (supertipo) que no pertenezcan a ninguno de los subtipos de entidad.
- *Con solapamiento* o *sin solapamiento*, dependiendo de si pueden existir o no ejemplares del supertipo de entidad que se relacionen con ejemplares de más de un subtipo distinto.

Pueden existir interrelaciones de cada una de las cuatro combinaciones posibles, y se representarían de la siguiente forma:

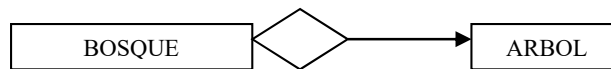


7. *Interrelaciones jerárquicas de agregación.* Permiten representar tipos de entidad compuestos que se forman como agregación de otros más simples. Se pueden clasificar a su vez en:

- *Agregación compuesto – componente.* El supertipo de entidad se obtiene por la unión de los subtipos. Se representa de la siguiente forma:



- *Agregación miembro – colección.* El supertipo de entidad es una colección de elementos de un mismo subtipo. Su representación sería así:

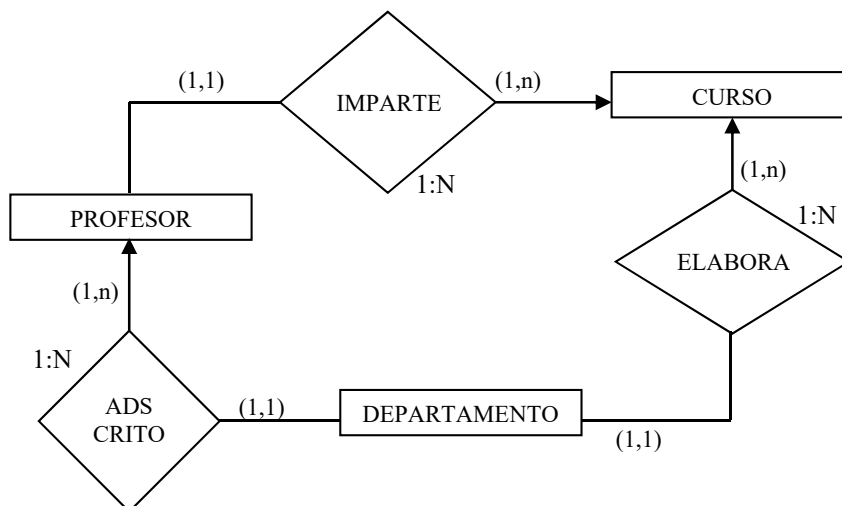


Normas para la creación del modelo E – R a partir del enunciado de un problema:

- Los *tipos de entidad* son objetos (sustantivos; nombres o complementos directos) que tienen por atributos a otros sustantivos además de su propio nombre. Se diferencian de los *atributos* en que suelen participar en más de un tipo de interrelación.
- Los *tipos de interrelación* son verbos que relacionan entre sí dos o más tipos de entidad (hacen posible la selección de ejemplares de un tipo de entidad refiriéndose a atributos de otros tipos de entidad).
- Los nombres propios suelen ser *ejemplares* de algún tipo de entidad.

Control de redundancia. Es aconsejable vigilar y eliminar las posibles redundancias en la creación del modelo. Pueden ser de dos tipos:

- Redundancia en *atributos*, cuando un atributo se puede calcular a partir de otro o de una combinación de varios.
- Redundancia en *interrelaciones*, cuando la existencia de un tipo de interrelación viene implícita por la existencia de otro u otros tipos de interrelación diferentes. Por ejemplo, en el caso siguiente, como un profesor pertenece a un único departamento y un curso lo imparte un único profesor, ya sabemos qué departamento ha elaborado el curso. Sobraría el tipo de interrelación *ELABORA* o, de la misma manera, el tipo de interrelación *ADSCRITO*.

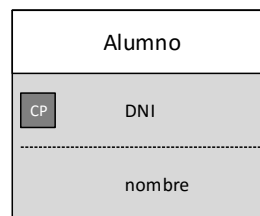


Existen otras notaciones utilizadas en el modelo ER, además de la propuesta por Chen. Las más extendidas son:

- Modelo IE (*Information Engineering*), también conocido como *Crow's Foot* por el símbolo utilizado para la cardinalidad n . Es un modelo sencillo y más aproximado al modelo lógico que el modelo original.
- Modelo IDEF1X (*Integrated Definition 1, Extended*). Más difícil de entender y de utilizar.
- Modelo UML. Utiliza características pensadas para la programación orientada a objetos.

Modelo IE. Esta notación se aproxima más a la utilizada en los modelos lógico y físico, y permite especificar detalles propios de estos modelos.

Las entidades en este modelo se muestran mediante un rectángulo con el nombre en la parte superior. En su interior pueden figurar los atributos de la clave primaria y, separados de éstos mediante una línea, el resto de atributos de la relación.

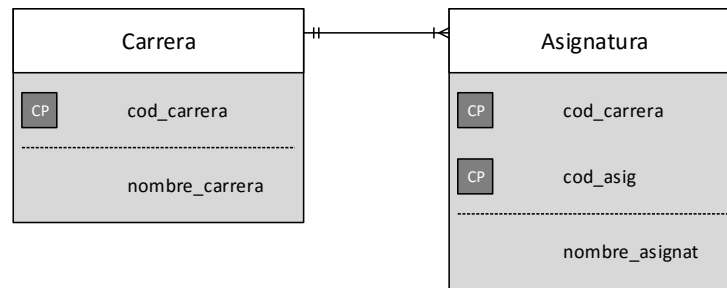


Las relaciones son líneas que unen entre sí las entidades. Las cardinalidades mínima y máxima se indican mediante dos símbolos en la línea; el símbolo más alejado de la entidad corresponde a la cardinalidad mínima, y el más próximo a la cardinalidad máxima.

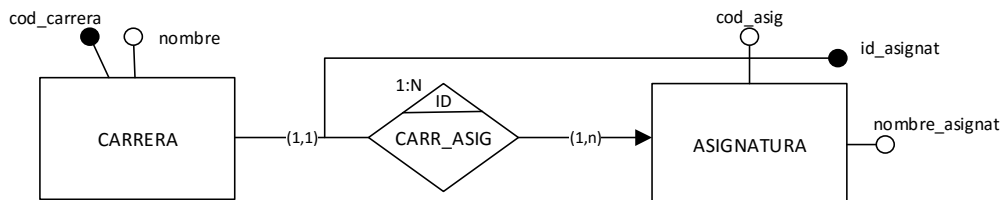
Estos símbolos pueden ser un círculo (0), una línea vertical (1) o un triángulo (n). Las distintas posibilidades son:

Símbolo	Significado
+○——	Cardinalidad mínima 0 y máxima 1.
⋈○——	Cardinalidad mínima 0 y máxima varios.
+ ——	Cardinalidad mínima 1 y máxima 1.
⋈ ——	Cardinalidad mínima 1 y máxima varios.

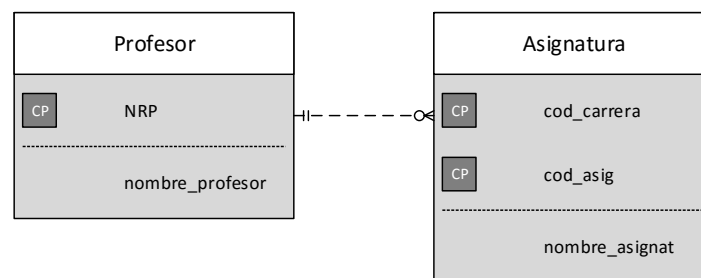
Además, si la línea es continua indica una dependencia en identificación de una entidad débil con respecto a otra fuerte.



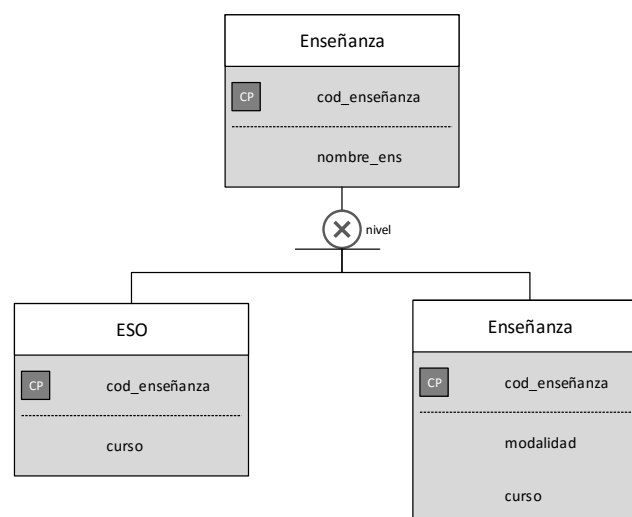
Que sería equivalente al siguiente modelo:



En algunos casos, además, la entidad débil en identificación se representa con los bordes redondeados. Si la relación entre las entidades no es débil en identificación, la línea debe ser discontinua.

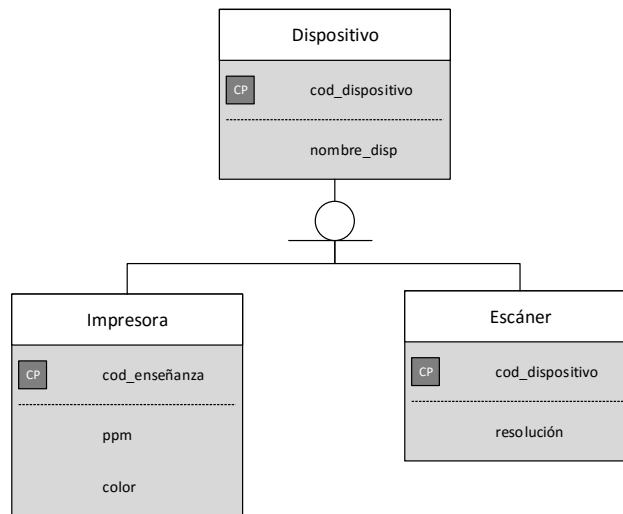


El modelo también permite especificar interrelaciones jerárquicas de generalización – especialización con la siguiente notación:



Si el círculo (que también se puede representar como un semicírculo) lleva una cruz, se trata de una clasificación sin solapamiento o exclusiva (las entidades del supertipo no pueden pertenecer a más de un subtipo); en este caso, se puede acompañar del atributo diferenciador, como figura en el ejemplo.

Las clasificaciones inclusivas son similares, pero en este caso el círculo o semicírculo no lleva cruz.



En este modelo, las clasificaciones no permiten especificar si son totales o parciales. Tampoco están soportadas por el modelo:

- relaciones de grado superior a 2
- atributos en las relaciones

Además, para facilitar el paso a los modelos lógico y físico, muchas herramientas de generación de diagramas no soportan relaciones de grado 2 con tipo de correspondencia N:M “muchos a muchos”. En todos estos casos, es necesario crear una nueva entidad intermedia.